Explaining Measurement Nonequivalence Using Multilevel Structural Equation Modeling: The Case of Attitudes Toward Citizenship Rights

Davidov, Eldad ; Dülmer, Hermann ; Cieciuch, Jan ; Kuntz, Anabel ; Seddig, Daniel ; Schmidt, Peter

Abstract: It is necessary to test for equivalence of measurements across groups to guarantee that comparisons of regression coefficients or mean scores of a latent factor are meaningful. Unfortunately, when tested, many scales display non-equivalence. Several researchers have suggested that non-equivalence may be used as a useful source of information as to why equivalence is biased and proposed employing a multilevel structural equation modeling (MLSEM) approach to explain why equivalence is not given. This method can consider a latent between-level factor and/or single contextual variables and use them to explain items' non-equivalence. In the current study we show that this method may also be useful for social science studies in general and for survey research and sociological comparative studies in particular when one fails to establish cross-group equivalence. We utilize data from the International Social Survey Program (ISSP) national identity module (2003) to test for the cross-country equivalence of a scale measuring attitudes toward granting citizenship rights to immigrants. As expected, the scale fails to achieve scalar equivalence. However, we explain a significant part of the most non-equivalent intercept by a latent between-level factor and one contextual variable, namely, the percentage of foreigners in the country relying on group threat theory. We show that the method does not necessarily rectify non-equivalence but it can help to explain why it is absent.

DOI: https://doi.org/10.1177/0049124116672678

Originally published at:
Davidov, Eldad; Dülmer, Hermann; Cieciuch, Jan; Kuntz, Anabel; Seddig, Daniel; Schmidt, Peter (2018). Explaining Measurement Nonequivalence Using Multilevel Structural Equation Modeling: The Case of Attitudes Toward Citizenship Rights. Sociological Methods Research, 47(4):729-760.
DOI: https://doi.org/10.1177/0049124116672678

Explaining Measurement Nonequivalence Using Multilevel Structural Equation Modeling : The Case of Attitudes Toward Citizenship Rights

Eldad Davidov (1,2,3)

Hermann Dülmer (4)

Jan Cieciuch (1,3,5)

Anabel Kuntz (1)

Daniel Seddig (1)

Peter Schmidt (6)

[1]University of Zurich, Zurich, Switzerland
[2]University of Cologne, Cologne, Germany
[3]University Research Priority Program "Social Networks," University of Zurich, Zurich, Switzerland
[4]University of Würzburg and University of Cologne, Germany
[5]Cardinal Stefan Wyszyński University, Warsaw, Poland
[6]University of Giessen, Giessen, Germany

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in ***Sociological Methods & Research*** following peer review. It was first published online in this journal on October 17, 2016.

The definitive publisher-authenticated version is available online at Sage via

http://dx.doi.org/10.1177/0049124116672678

Abstract

It is necessary to test for equivalence of measurements across groups to guarantee that comparisons of regression coefficients or mean scores of a latent factor are meaningful. Unfortunately, when tested, many scales display non-equivalence. Several researchers have suggested that non-equivalence may be used as a useful source of information as to why equivalence is biased and proposed employing a multilevel structural equation modeling (MLSEM) approach to *explain* why equivalence is not given. This method can consider a latent between-level factor and/or single contextual variables and use them to explain items' non-equivalence. In the current study we show that this method may also be useful for social science studies in general and for survey research and sociological comparative studies in particular when one fails to establish cross-group equivalence. We utilize data from the International Social Survey Program (ISSP) national identity module (2003) to test for the cross-country equivalence of a scale measuring attitudes toward granting citizenship rights to immigrants. As expected, the scale fails to achieve scalar equivalence. However, we explain a significant part of the most non-equivalent intercept by a latent between-level factor and one contextual variable, namely, the percentage of foreigners in the country relying on group threat theory. We show that the method does not necessarily rectify non-equivalence but it can help to explain why it is absent.

Key words: measurement equivalence; multilevel structural equation modeling; ISSP; percentage of foreigners in the country; attitudes toward granting citizenship rights to immigrants

Introduction

Comparative sociology requires that concepts are equivalent (i.e., measurement invariant) across countries, cultures, regions, or time points[1]. Various statistical procedures have been developed to test for measurement equivalence. Yet such statistical analyses often suggest that concepts are not comparable after all (see, e.g., Ariely and Davidov 2011; Billiet 2003; Meuleman, Davidov, and Billiet 2009; Davidov 2008, 2009; Davidov, Schmidt, and Schwartz 2008; De Beuckelaer, Lievens, and Swinnen 2007; Van der Veld and Saris 2011). As a result, researchers are confronted with the question of what useful information their data can provide even though it lacks equivalence across their groups under study (Byrne and van de Vijver 2010; Fischer, Milfont, and Gouveia 2011; Fontaine et al. 2008). This problem is particularly troublesome nowadays given the abundance of multicountry and longitudinal surveys. These surveys have the potential to provide useful information about similarities and differences across countries and over time. However, they also involve the risk that such comparisons may be problematic when equivalence is not given. On the one hand, if researchers ignore the finding that concepts are not comparable and continue with substantive comparative work, it could well be the case that findings are biased and conclusions may not be meaningful, as several previous studies have demonstrated (see, e.g., Kuha and Moustaki 2015). On the other hand, if researchers decide in such a case to refrain from any comparisons, it may have the consequence that data are not exhausted and their potential is not realized. Therefore, several methodologists have proposed that when equivalence is not established across groups, researchers may try to *explain* why it is not given (Davidov et al. 2014).

In the current study we are going to present for this purpose the use of multilevel structural equation modeling (MLSEM). We will demonstrate that the method can also be applied for explaining non-equivalence of sociological or other social scientific concepts in a theoretically driven way. The methodology and technique we are going to present and apply

[1] In the current study we will use the concepts invariance or equivalence interchangeably. Both are commonly used in the literature.

in the study are not new. They have been reported in previous (recent and older) studies (see, e.g., Cheung and Au 2005; Davidov et al. 2012; Jak et al. 2014a, b; Hox 2010; Muthén 1989, 1994, 2011; Rabe-Hesketh et al. 2004). However, even though several researchers have underlined the importance of explaining non-invariance rather than only testing for it, we have not seen any application of this method in comparative studies in the social sciences (other than psychology), where 'real' survey data from large population studies are used and sociological contextual variables are applied to explain non-invariance (for an exception in the psychological literature, see Davidov et al. 2012). After all, many invariance tests fail to establish full or even partial invariance. Therefore, we find it extremely important to also introduce this method to survey researchers and comparative social scientists and provide an example of its practical application on theory and measurements which are more appealing to this readership.

The study will focus on a scale from the International Social Survey Program (ISSP) national identity module to measure attitudes toward granting citizenship rights to immigrants. This scale has been often employed but its equivalence properties have not been tested across a large set of countries. This has been the case for most other scales in the ISSP as well (for a few exceptions, some of which focus on a small set of countries or on other scales in the ISSP, see Braun, Behr, and Kaczmirek 2013; Davidov 2009, 2011; Raijman et al. 2008; Rammstedt, Goldberg, and Borg 2010; Reeskens and Hooghe 2010). In the current study we will test its equivalence across the ISSP countries and try to explain the absence of equivalence in one of its items in a theoretically driven way.

We proceed in the following way. First, in the *Measurement invariance* section, we will describe what measurement invariance is and explain how it can be tested using multigroup confirmatory factor analysis (MGCFA). Second, in the section *What can be done when measurement invariance is not given?,* we will list strategies for dealing with a situation in which measurement invariance is not given. Third, in the section *Using multilevel* *structural equation modeling to explain non-invariance,* we will present the method of multilevel structural equation modeling (MLSEM) and show how it may be used to explain non-equivalence. Fourth, in the *Theory* section, we will provide some theoretical mechanisms leaning on group threat theory for explaining why measurement invariance may be absent and formulate an empirically testable hypothesis for the explanation. Fifth, in the *Data and measurements* section, we will present the ISSP data we are using to illustrate this procedure as well as the individual and country-level measurements included in the analyses. In particular, we will describe the measurement of the concept of willingness to grant citizenship rights. In the *Statistical analysis* section we will test the invariance properties of the scale and then use MLSEM to test our hypothesis and explain the cross-country non-equivalence of one of the scale's items. We will close with a summary which includes a discussion of the limitations of the MLSEM method for explaining non-equivalence as well as with some concluding remarks.

Measurement invariance

Measurement invariance can be defined as "a property of a measurement instrument (in the case of survey research, a questionnaire), implying that the instrument measures the same concept in the same way across various subgroups of respondents" (Davidov et al. 2014:58). Horn and McArdle (1992:117) define measurement invariance as "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute". Thus, testing for invariance of measurements across countries and over time is necessary before meaningful comparisons of relationships and means may be conducted (Billiet 2003). Indeed, measurement invariance is a necessary condition that ensures that question items are perceived in a similar way and may be compared (Byrne and van de Vijver 2010). Because measurement invariance is a precondition of meaningful comparisons across groups, it is recommended to test it empirically before conducting the substantive analysis. There are several methods to test for measurement invariance, and among them, multigroup confirmatory factor analysis (MGCFA, Jöreskog 1971; Bollen 1989) is most often used. This method involves setting cross-group constraints on parameters and comparing more restricted models with less restricted ones (Steenkamp and Baumgartner 1998; Vandenberg and Lance, 2000, but for other methods see, e.g., Davidov, Schmidt, and Billiet 2011).

Researchers usually distinguish three basic levels of measurement invariance: configural, metric, and scalar (Vandenberg and Lance 2000), and each level is defined by the parameters constrained to be equal across groups. Configural invariance requires that identical latent variables are measured by the same items in all groups. Metric invariance requires that loadings of items on the latent variable are equal across groups. Scalar invariance means that both factor loadings and indicator intercepts are equal across groups (Vandenberg and Lance 2000). The configural level of measurement invariance ensures that the structure is similar across groups. Metric invariance justifies the assumption that the meaning of the latent variable in all groups is the same and allows correlates of the measured variable (covariances, unstandardized regression coefficients) to be compared across groups. Scalar invariance indicates that respondents use the scale in the same way in each group and allows the latent means to be compared across groups (Davidov et al. 2014; Vandenberg and Lance 2000).

The model fit of the configural level is the baseline against which the more constrained models are compared and analyzed. To assess whether a given level of measurement invariance is established, model fit measures are compared between the more and less constrained models (Chen 2007). Whereas lower levels (i.e., configural or metric) of invariance are often supported by the data in cross-national studies, this becomes increasingly seldom when higher levels (i.e., scalar) of invariance are tested across cultures or countries. Indeed, scalar non-invariance constitutes one of the most serious threats to cross-cultural research, and it is also the focus of the present study.

What can be done when measurement invariance is not given?

The question arises as to what can be done when measurement invariance is not given. Up to now, two solutions have been used most often in the literature (Davidov et al. 2012): The first one suggests eliminating non-invariance and the second one suggests liberalizing the criteria that provide evidence for invariance. Following the first solution it is possible to drop from further analysis those groups where non-invariance was identified or to release constraints on items which were identified as non-invariant. Unfortunately, both possibilities have disadvantages. Dropping groups implies losing data and information; therefore, this procedure may affect the substantive conclusions and limit further research. Releasing constraints on specific items builds on the concept of partial measurement invariance (Byrne, Shavelson, and Muthén 1989; Steenkamp and Baumgartner 1998). Partial invariance is supported when the parameters of at least two indicators per construct (i.e., loadings for partial metric invariance and loadings plus intercepts for partial scalar invariance) are equal across groups. According to Byrne et al. (1989) and Steenkamp and Baumgartner (1998), partial invariance is sufficient for meaningful cross-group comparison. However, more recent studies question whether partial invariance may indeed be sufficient (de Beuckelaer and Swinnen 2011; Steinmetz 2011).

The approach which suggests liberalizing the criteria that provide evidence for invariance was recently proposed by Muthén and Asparouhov (2013) who argued that the commonly used criteria for evaluating measurement invariance are too strict. Instead, they proposed testing for approximate measurement invariance (see also van de Schoot et al. 2013). Approximate measurement invariance allows small differences in loadings or intercepts across groups. Moreover, testing for approximate measurement invariance can be performed within the Bayesian SEM framework (Cieciuch et al. 2014; Muthén and Asparouhov 2013; van de Schoot et al. 2013; Zercher et al. 2015).

Both solutions discussed above assume that non-invariance prevents researchers from carrying on with conducting a meaningful, theoretically driven analysis and from drawing meaningful conclusions about cross-cultural similarities or differences. We propose to reconceptualize the problem and investigate *why* invariance is not found. After all, measurement non-invariance can provide interesting information about differences or similarities across groups. In the next section we will show how (MLSEM) can be used to explain measurement non-invariance.

Using multilevel structural equation modeling to explain non-invariance

Although MLSEM was introduced some two decades ago (cf. Cheung and Au 2005; Hox 2010; Muthén 1985, 1994), its further development has been the focus of more recent empirical investigation (Cheung and Au 2005; Davidov et al. 2012; Jak et al. 2014a, b; Hox 2010; Muthén 1989, 1994, 2011; Rabe-Hesketh et al. 2004). However, its application has become more accessible to applied researchers in recent years only after its inclusion in structural equation modeling software packages such as M*plus* (Muthén and Muthén 1998-2014). To the best of our knowledge, it has not yet been applied in sociology. The basic concept of MLSEM, like in multilevel regression models, is to decompose the variability of the indicators into individual ("within") and contextual ("between"; e.g., country) variability.

In order to use MLSEM techniques to explain scalar measurement non-invariance, we proceed in four steps. In the first step we propose to test for measurement invariance using an MGCFA. The advantage of first using an MGCFA is that it allows testing systematically for different levels of measurement invariance. In the second step we propose fitting the data in a multilevel CFA. Metric invariance and absent scalar invariance are reflected in a multilevel CFA by a significant unexplained between-level variance of at least one item's error term. A significant between-level error term implies that the intercept of the corresponding item is considerably different across groups (Davidov et al. 2012). In the third step we propose testing whether a between-level latent variable may already account at least partially for the between-level error variance of an item. Finally, in the fourth step, we propose adding, in a theoretically driven way, contextual predictors of the between-level error term to explain scalar non-invariance of the corresponding item.

Figure 1a corresponds with step 3 and illustrates an MLCFA with one latent factor at Level 1 (within) and one latent factor at Level 2 (between) with k = 3 Level 1 indicator variables. The corresponding general formula for a two-level CFA (cf. also Muthén 1991:344) is given below the figure.

Figure 1a about here

Level 1 (within): Level 2 (between):

$$y_{kij} = \alpha_{kj} + \lambda_{Wk} \cdot \eta_{Wij} + \varepsilon_{Wkij} \qquad \alpha_{kj} = \upsilon_k + \lambda_{Bk} \cdot \eta_{Bj} + \varepsilon_{Bkj} \tag{1}$$

where

- y_{kij} refers to the observed value of respondent i of country j on indicator variable k,
- α_{kj} refers to the intercept of indicator variable k in country j,
- υ_k refers to the cross-country grand intercept of indicator variable k (i.e., the grand mean when the between-level latent variable equals zero),
- η_{Wij} refers to the score of respondent i of country j on the within-level latent variable η_W,
- η_{Bj} refers to the score of country j on the between-level latent variable η_B,
- λ_{Wk} refers to the within-level factor loading λ_W of indicator variable k,
- λ_{Bk} refers to the between-level factor loading λ_B of indicator variable k,

- ε_{Wkij} refers to the within-level error term ε_W for respondent *i* of country *j* on indicator variable *k*, and
- ε_{Bkj} refers to the between-level error term ε_B (usually called random intercept term in multilevel analysis) for country *j* on indicator variable *k*.

The above outlined two-level CFA model can generally be affected by measurement non-invariance in various ways (for a systematic elaboration of the connection between the multilevel approach and measurement invariance, see Fontaine 2008; Jak, Oort, and Dolan 2013, 2014a, b). Unequal factor loadings across groups (*deviations from metric invariance*) can be modeled by allowing one or more random slopes for the within-level factor loadings (Chen, Bliese, and Mattieu 2005; Kim et al. 2015; Schlüter and Meuleman 2009). Cross-group intercept differences (*deviations from scalar invariance when metric invariance holds*) find their expression in the between-level error terms ε_{Bkj}.[2] In other words, substantial *between-level error variance in the indicators* as captured in the multilevel analysis by the random components implies *unequal item intercepts* or deviations from scalar equivalence for models with equal factor loadings across groups (Davidov et al. 2012; Kim et al. 2015).

Furthermore, several authors have argued that, to perform meaningful MLSEM, specific assumptions have to be made about measurement invariance. Cheung, Leung, and Au (2006:523), for example, stress that the within-factor structure should hold across groups and propose to test this assumption with meta-analytic structural equation modeling (MASEM). Fontaine (2008:77-78) similarly emphasizes that relations between latent factors and indicators should be identical (or very similar) across groups and that the country-level error

[2] The size of the between-level error variance in the indicators cannot be interpreted when metric invariance fails and within factor loadings have to be estimated with a random component (cf. Hox 2010:60-61).

terms should be (very close to) zero.[3] Others suggest that the loadings of corresponding items should be equal across levels (Jak, Oort, and Dolan 2014a).

In this study, we argue that in order to draw meaningful conclusions from MLSEM, equal within-level factor loadings and item intercepts are required (Steenkamp and Baumgartner 1998) (but not necessarily equal loadings across levels; see Kim et al. 2015, p. 606). When these assumptions are not met, correcting for measurement non-invariance or explaining it at least partially is a sensible option (Fontaine 2008:78).

A further "correction" or explanation of non-invariance is subject of the fourth and last step of the procedure. Cross-group differences in the parameters (e.g., intercepts) are taken into account by including single (individual and/or) contextual predictors in the model (see Jak 2014; Jak et al. 2013, 2014a, b), hence *explaining* scalar measurement non-invariance. This approach, which is an extension of MLCFA (cf. Hox 2010; Muthén 1994) to MLSEM (cf. Muthén 1994; Selig, Card, and Little 2008), is not an alternative to the cross-cultural comparison of the theoretical concepts of interest. Rather, it constitutes an additional useful test to explain why invariance does not hold. More specifically, contextual predictors are included in order to further explain Level 2 variability of the indicator (α_{kj}) and to reduce the unexplained country-level variance of the indicators intercepts (ε_{Bkj}). By completely reducing the remaining variability in the intercept, the between-level error term ε_{Bkj} should approach zero, or in other words, be fully accounted for. However, even a partial explanation of the variability in the intercepts is a viable option, especially if it is a result of a theoretically driven explanation. Assuming countries as the context, country characteristics included as Level 2 predictors could be aggregates of individual-level variables such as employment status, average education, average level of religiosity, and average income, or variables that

[3] Although the arguments of Cheung et al. (2006) and Fontaine (2008) refer to equivalence on the within level, they are not identical. While Cheung et al. (2006) discuss equivalence in terms of the homogeneity of correlation matrices (equal factor loadings, error covariances, and factor (co)variances across groups), the argument developed by Fontaine (2008), on the other hand, also considers the mean structure of the data (additional equivalence of the item intercepts across countries).

characterize the country level such as the number of immigrants in a country, policies, history, inflation rate, and other economic conditions.

In Figure 1b, which corresponds to step 4, the above introduced MLCFA is extended to an MLSEM by including a country-level predictor variable, and this is typically called a z-variable in multilevel analyses. The corresponding general formula for a two-level SEM with one latent variable at each level is given below the figure.

Figure 1b about here

General formula for a 2-level SEM with one latent variable at each level:

Level 1 (within):

$$y_{kij} = \alpha_{kj} + \lambda_{Wk} \cdot \eta_{Wij} + \varepsilon_{Wkij}$$

Level 2 (between):

$$\alpha_{kj} = \upsilon_k + \lambda_{Bk} \cdot \eta_{Bj} + \sum_{l=1}^{L} \sum_{k=1}^{K} \gamma_{lk} \cdot z_{lj} + \varepsilon_{Bkj} \quad (2)$$

$$\eta_{Bj} = \mu_B + \sum_{l=1}^{L} \gamma_{l0} \cdot z_{lj} + \zeta_{Bj}$$

where

- z_{lj} refers to the observed value of country *j* on a country-level predictor variable z_l,
- γ_{l0} refers to the regression coefficient of the latent variable η_{Bj} on a country-level predictor variable z_{lj},
- γ_{lk} refers to the regression coefficient of α_{kj} of an indicator variable k on a country-level predictor variable z_{lj},
- μ_B refers to the intercept of the regression from the latent variable η_{Bj} on the country-level predictor variable z_{lj}, and

- ζ_{Bj} refers to the between-level residual error term of the latent variable η_{Bj} (the remaining abbreviations are identical to those used for Formula 1).

In the following, we will illustrate how the method may be used to explain scalar non-invariance for the measurement of respondents' willingness to concede citizenship rights to immigrants as measured in the ISSP national identity module data. The present demonstration will show first how to model a two-level CFA. Then, by including a context-level predictor and extending the two-level CFA into a two-level SEM, it will illustrate how scalar non-invariance of one of the items is significantly accounted for in this way. The following theoretical discussion is presented to justify the use of a specific country-level variable, the size of the immigrant population in a country, to explain scalar non-invariance.

Theory

As countries' citizenship rights are commonly conceived of as a means for social closure (Levanon and Lewin-Epstein 2010; Raijman et al. 2008), people's willingness to concede citizenship rights equals their readiness to exclude immigrants from the national ingroup. Citizenship rights regulate the incorporation of immigrants into the society by defining their legal status, their political participation, various societal rights and obligations as well as integration into national identity (Ceobanu and Escandell 2011). By expressing exclusionary attitudes, individuals thereby support to restrict immigrants' access to the national ingroups' rights and privileges (e.g., Gorodzeisky and Semyonov 2009; Levanon and Lewin-Epstein 2010; Raijman et al. 2008). Attitudes toward citizenship rights can be described along the two predominant principles according to which countries legally grant citizenship rights to individuals, namely, by territory and by descent. When based on descent, citizenship rights are exclusively granted to individuals with national ancestry (*jus sanguinis*). When based on territory, citizenship rights are granted to individuals who were either born in the country (*jus* *solis*) or who have lived in the country for a certain time period (*jus domicile*) (Ceobanu and Escandell 2011; Hochman, Raijman and Schmidt 2015; Levanon and Lewin-Epstein 2010; Raijman et al. 2008). These non-mutually exclusive citizenship rights principles can be placed on a continuum from most exclusionist to most inclusive: Whereas citizenship rights based on ancestry (jus sanguinis) are most exclusive, citizenship rights based on territory indicate inclusiveness and immigrant integration (Ceobanu and Escandell 2011).

With rising levels of ethnic diversity in modern societies, people's willingness to include immigrants into the national community and thereby their willingness to share the national benefits might decline (e.g., Gorodzeisky and Semyonov 2009; Levanon and Lewin-Epstein 2010; Wright 2011). Accordingly, one prominent line in cross-national research on exclusionary attitudes toward immigrants, group threat theory, has focused on the share of immigrants in a country to explain these attitudes (e.g., Blalock 1967; Blumer 1958; Quillian 1995; Gorodzeisky and Semyonov 2009; Scheepers, Gijsberts, and Coenders 2002; Schlüter and Scheepers 2010; Schneider 2008). By regarding the size of the immigrant population as a dominant source of intergroup competition, group threat theory contends that actual competition over scarce resources leads to increased perceptions of threat. These perceptions of threat and competition in turn lead to exclusionary attitudes toward immigrants. Negative attitudes and exclusionary attitudes toward immigrants serve to maintain the group's status, its exclusive access to resources, privileges, and rights as well as its culture (e.g., Blumer 1958; Gorodzeisky and Semyonov 2009; Scheepers et al. 2002; Schlüter and Scheepers 2010; Schneider 2008). Depending on individuals' predispositions and characteristics, peoples' level of perceived threat to self-interest or collective interests might differ. Nevertheless, independent of perceived threats to self-interest, individuals might still perceive group interests to be at stake (e.g., Schlüter and Scheepers 2010). Indeed, research on exclusionary attitudes toward citizenship rights conceives these attitudes as a result of perceived threat to the national and cultural cohesion and community. Restricting immigrants' membership by expressing exclusionary attitudes to citizenship rights can, therefore, be assumed to function as means to restrict immigrants' access to the rights and privileges of the national ingroups and to preserve national, cultural, and social cohesion (e.g., Ceobanu and Escandell 2011; Gorodzeisky and Semyonov 2009; Levanon and Lewin-Epstein 2010). Therefore, we postulate the following hypothesis between the contextual independent (macro) variable and a dependent latent (macro) variable:

H1 The larger the immigrant population rate in a country, the lower the publics' willingness to concede citizenship rights to immigrants.

As stated above, the willingness to grant citizenship rights to immigrants differs on the exclusive-inclusive dimension. Likewise, Raijman and colleagues contend granting citizenship rights to immigrants to be a crucial test of a liberal attitude to foreigners' incorporation into the host society (Raijman et al. 2008:204). At the same time, willingness to grant citizenship rights to immigrants indicates a more inclusive attitude than willingness to grant citizenship rights to foreigners' *children* born and socialized in the country. Indeed, as countries with larger immigrant populations are expected to be more threatened due to immigration (than due to immigrants' children), incorporating legal immigrants into the national community can be assumed as a larger threat than incorporating the children of immigrants into the national community. Therefore, we propose the following hypothesis between the contextual independent (macro) variable and the dependent (macro) observed variable:

H2 The larger the immigrant population rate in a country, the lower the level of willingness to concede equal rights to legal immigrants (as compared to their offspring).

As will be shown below, two items in our scale measuring willingness to grant citizenship rights refer to immigrants' children whereas only the third item refers to immigrants per se. Based on the explanations above we expect this item to behave differently than the other two items, and we assume that its specific cross-country variation may be accounted for, at least partly, by the relative size of the immigrant population in a country. Next, we turn to the empirical test where we first assess the invariance properties of the scale measuring attitudes toward granting citizenship rights to immigrants by an MGCFA analysis and then try to explain the non-invariance in this specific item.

Data and measurements

Individual-level data comes from the National Identity Module 2003 of the International Social Survey Programme (ISSP Research Group 2012) covering respondents from various countries around the globe.[4] Data of the national representative adult population was collected by means of face-to-face interviews or self-completion questionnaires. The fieldwork was mostly carried out in 2003 and/or 2004.[5] The *willingness to concede citizenship rights to immigrants* is assessed by three statements, each reflecting one of the main citizenship principles: Support of the jus sanguinis principle is measured by the statement (1) "Children born abroad should have the right to become COUNTRY NATIONALITY citizens if at least one of their parents is a COUNTRY NATIONALITY citizen." Support for the jus solis principle is measured by the statement (2) "Children born in COUNTRY of parents who are not citizens should have the right to become COUNTRY NATIONALITY citizens." Support for the jus domicile principle is measured by the statement (3) "Legal immigrants to COUNTRY who are not citizens should have the same rights as COUNTRY NATIONALITY citizens."[6] Respondents indicated their support to each of the statements on a 5-point Likert-type scale ranging from 1 (agree strongly) to 5 (disagree strongly). For easier interpretation

---

[4] We treated East and West Germany as well as Israeli Arabs and Israeli Jews separately in our analysis. Due to missing information on the contextual level, Taiwan had to be excluded from the analysis (cf. the Appendix). Thus, a total number of 35 groups was included in the analyses.

[5] Further details on data collection and documentation can be downloaded from http://www.gesis.org/issp/issp-modules-profiles/national-identity/2003/; the data can be downloaded at no cost from http://zacat.gesis.org/webview/index.jsp?object=http://zacat.gesis.org/obj/fStudy/ZA3910

[6] Note that this is the only item which refers to immigrants themselves rather than to their children.

we reversed the answer scale and subtracted one unit (i.e., the new scale ranged from 0 “disagree strongly” to 4 “agree strongly”).

The *size of the immigrant population* was assessed using the average percentage of the foreign born population in a country (including refugees) from 1995 to 1999 (The World Bank Group 2014). The respective percentages for each country included in our analyses are documented in the Appendix.

Statistical analyses

Descriptive statistics

Reflective indicators that belong to the same latent variable are expected to be highly correlated (Byrne 2010). To test this expectation we computed the observed correlations and covariances of the three indicators. The results are reported in Table 1. The correlations on the individual level ranged between 0.265 and 0.378. The correlations on the country level were somewhat stronger, ranging between 0.344 and 0.633. Altogether, the results confirmed our expectation of sufficiently strong correlations thus allowing us to conduct a CFA for the three indicator variables on both levels.

Table 1 about here

Testing for invariance (Step 1)

We tested the measurement invariance properties of the scale across countries by comparing the global fit measures of higher levels of invariance (with more restricted parameters) to those of lower levels of invariance (with fewer restricted parameters). We also inspected the change in the global fit indices between the measurement invariance levels. The results are presented in Table 2.

Table 2 about here

To evaluate the global fit indices we followed the recommendations proposed by Hu and Bentler (1999), Marsh, Hau, and Wen (2004), and West, Taylor, and Wu (2012): Specifically, we considered comparative fit index values (CFI) greater than .90, root mean square error of approximation (RMSEA) values less than .08, and root mean square residual (SRMR) values less than .08 as indicators of reasonable model fit and CFI values greater than .95, RMSEA values less than .05, and SRMR values less than .05 as indicators of very good model fit. The configural level was just identified because in the case of only three indicators, the number of degrees of freedom equals zero. As Table 2 demonstrates, according to the cut-off values outlined above, the global fit measures of the metric invariance model fit the data very well.

To evaluate the change in the global fit indices between the metric and scalar levels we followed Chen's (2007) recommendations for samples larger than 300. According to Chen, a change in CFI less than .01, a change in SRMR less than .01, and a change in RMSEA less than .015 when moving from metric to scalar level of measurement invariance indicate the presence of scalar measurement invariance. Because the observed changes for CFI (Δ = .352), SRMR (Δ = .078), and RMSEA (Δ = .108) undoubtedly exceeded the recommended cut-off criteria, we ascertained the lack of scalar measurement invariance of respondents' willingness to concede citizenship rights to immigrants across ISSP countries. In addition, we inspected the size of local misspecifications in the scalar invariance model using the software package Jrule (Oberski 2009) according to the recommendations of Saris, Satorra, and van der Veld (2009). These findings revealed that the intercept of the three items in the scalar invariance model was misspecified in most countries thus supporting the conclusion that the model lacks scalar measurement invariance.[7]

[7] It should be noted that whereas the MGCFA tells how many countries significantly deviate from the common intercept, MLCFA tells us how large the (squared) sum (variance components) of the deviations from the common intercept is, independent of how many countries contribute to scalar measurement non-invariance.

Multilevel CFA and multilevel SEM (Steps 2 – 4)

In this part of the analysis we tried to explain the non-invariance observed in the previous section. First, we estimated a two-level model for which a latent factor was only estimated at Level 1, whereby the indicator variables measuring willingness to grant citizenship rights were allowed to have between-level variability (Step 2, see Figure 2a). Then we extended the model to a full two-level CFA by including a latent factor at the between-level (Step 3, see Figure 2b). Finally, we regressed the between-level latent variable as well as the jus domicile item (i.e., the item's α_j) on the percentage of foreign-born population of a country to test Hypotheses 1 and 2 (Step 4, see Figure 2c). Table 3 presents the results of these three models. The graphical representation of the results is displayed in Figures 2a to 2c.

Table 3 and Figures 2a to 2c about here

The global fit measures of Model 1 (model without a level-2 latent factor) displayed significant between-level variances (random components) for all three indicator variables, confirming that scalar invariance was absent for all of them. The jus domicile item, which measured the willingness to concede citizenship rights to legal immigrants, turned out to have the largest country-level variability (0.111), followed by the jus solis (0.077), and jus sanguinis (0.045) items. This model did not have satisfactory model fit as indicated by the SRMR global fit measure for the between-level part of the model (0.368). Introducing the between-level latent factor for citizenship, as was done for Model 2 (2-level CFA), led to an acceptable model fit.[8] The between-level residual variance (random component) of the jus

[8] Some authors (cf., for instance, Jak et al. 2014b) suggest that fitting a MLCFA that represents metric invariance across groups requires imposing an equality constraint on the factor loadings of corresponding items across levels. We think that this requirement is too strict for our purposes because it tests for a different and more restrictive model than we actually need (see also Kim et al. 2015; see also Chen et al. 2005). In addition, it does also not hold for a lot of applications in sociology. Since the Level 1 factor loadings of jus sanguinis and jus domicile (0.709 and 0.782) in our model differ too much from the respective factor loadings of Level 2 (1.198 and 1.026), no equality constraint has been imposed in the MLCFA. The substantive conclusions remain,

sanguinis item became non-significant ($p = 0.996$) and was fixed for this reason to zero. The between-level residual variances of the other two indicators also decreased: The random component of jus solis decreased by 39 percent from 0.077 to 0.047, and the random component of jus domicile decreased by approximately 29 percent from 0.111 to 0.079.

The results of the MLSEM model (Model 3) are depicted in Figure 2c. In this model, we regressed the latent variable measuring the willingness to concede citizenship rights to non-citizens on the country-level variable measuring the share of the foreign-born population in the country. In addition, we regressed the between-level item measuring the willingness to concede citizenship rights to legal immigrants (the jus domicile item) on the same country-level variable. The results provided support for Hypothesis 1: The higher the percentage of the foreign-born population in a country, the lower was the publics' willingness to concede citizenship rights to immigrants (b = -0.007, z = -1.998, beta = -0.345). In addition, this effect is significant at the 5% level (one-tailed test of significance). Hypothesis 2 was also empirically supported by our data: The higher the percentage of the foreign-born population in a country, the lower was the willingness in a country to concede citizenship rights to legal immigrants who are not citizens of the country (b = -0.015, z = -2.951, beta = -0.391). Thus, the percentage of the foreign-born population in a country contributed significantly to explain why scalar non-invariance was not evidenced for the jus domicile item. By regressing the between-level indicator of jus domicile on the percentage of a country's foreign-born population, the between-level residual variance (random component) of this item decreased by nearly 19 percent from 0.079 in Model 2 to 0.064 in Model 3.[9] Hence, country level differences in the intercept of the jus domicile item could be traced back to substantial differences in the percentage of the foreign-born population between countries.

---

however, in our case unchanged even when similarity constraints on the loadings across levels are imposed (output demonstrating this result may be received from the second author upon request).

[9] In addition, the Akaike information criterion (AIC; Bollen 1989) was lowest for Model 3. Since multilevel data have a different sample size at different levels, the interpretation of the AIC is more straightforward than that of the Bayesian information criterion (BIC; Bollen 1989) and, therefore, the recommended choice (cf. Hox 2010:51).

Summary and conclusions

Comparative sociology requires that the concepts under study are equivalent. However, it is often the case that statistical analyses fail to establish equivalence, particularly scalar equivalence, across the groups studied (i.e., countries, cultures, or time points). Failing to establish equivalence is a serious problem, because comparative analysis that ignores the absence of measurement invariance runs the risk of drawing wrong conclusions (Davidov et al. 2014; Kuha and Moustaki 2015; Steenkamp and Baumgartner 1998). Computation of incorrect group means may result in severely biased group rankings (Little 2013; Little, Lindenberger, and Nesselroade 1999; Steinmetz 2011, 2013), and computing erroneous associations in different groups may result in wrong conclusions about relations between variables of interest. This problem is particularly worrisome nowadays when the number of international and longitudinal surveys and the potential for performing comparative studies are higher than ever. Instead, several researchers have suggested that measurement non-invariance may also be considered as a useful source of information as to *why* invariance is not given. Recent methodological literature has shown how multilevel SEM may be used systematically to understand why invariance is absent. This represents a more sociological approach toward explaining scalar non-invariance in the sense of an extended measurement theory (Bollen 1989; Hempel 1952) than the classical differential item functioning analysis using only individual attributes like gender or age as predictors of item bias (Lee, Little, and Preacher 2011)[10]. MLSEM allows researchers to consider context-level variables that may explain, ideally in a theoretically driven way, why the parameters of a certain item (e.g., the intercept) vary across groups. As such, it can be used as a valuable tool to *explain* scalar non-

[10] A major difference between MLSEM and the detection of differential item bias (see, e.g., Lee et al. 2011; Muthén, Kao, and Burstein 1991; Thompson and Green 2013; Woods 2009) is that the former allows for explanations of item bias both on the context and individual levels of analysis whereas the latter allows for such explanations particularly on the individual level (Bollen 1989).

invariance. This is of particular relevance because it enables researchers to draw upon sociological theories to understand why measurements are not comparable across countries.

In this study we presented the MLSEM method to a sociological audience and demonstrated that the method can also be used to address and explain non-equivalence and item bias of sociological concepts in a theoretically driven way. The methodology and technique we presented and applied in this study are not new. However, we are not aware of any publications applying MLSEM in sociology to explain non-invariance, even though several researchers have recently underlined the importance of explaining non-invariance rather than only testing for it. This neglect is unfortunate because MLSEM may prove itself to be a powerful tool to explore and better understand the groups under study and potential differences in their measurement properties and response patterns. We began by explaining what measurement invariance is and how it may be tested using MGCFA. Then we presented the MLSEM method and clarified how it may be used to explain non-invariance.

For the empirical examination we used a scale from the ISSP national identity module from 2003 to measure attitudes toward granting citizenship rights to immigrants. This scale has been often employed[11], but its invariance properties have not been tested across a large set of countries. The invariance test revealed that whereas metric invariance across countries was supported by the data, scalar invariance was not supported by the data. It was notably absent in one of the three items measuring the agreement with the statement that legal immigrants who are not citizens should have the same rights as country citizens. MLSEM was performed to try to explain scalar non-invariance in this item. Our analyses were conducted in four steps: Whereas the first step was the MGCFA to test for measurement (non)invariance, in the second and third steps we employed MLCFA. In the former we did not include a between-level latent variable, and in the latter we added it to the model. This model accounted for variations in the indicators across individuals and countries by using individual and contextual

[11] For examples, see http://www.issp.org/uploads/editor_uploads/files/ISSP_2015Biblio.doc

latent variables. A final, explicitly theoretically driven explanation of non-invariance was performed in the fourth step. We introduced, on the country level, an independent contextual variable measuring the percentage of people born in a different country. This measure has often been used as an indication of threat due to immigrants (see, e.g., Semyonov, Raijman, and Gorodzeisky 2006; some authors refer to it as threat to self-interests, see e.g., Scheepers et al. 2002; Schneider 2008; other authors refer to it as a measure of threat to the national cohesion, the culture, and the community, see, e.g., Ceobanu and Escandell 2011; Gorodzeisky and Semyonov 2009; Schlüter and Scheepers 2010). Using this contextual measure and relying on group threat theory (Quillian 1995), we successfully explained part of the cross-country variation in the item intercept. In contrast to the publications mentioned above, we demonstrated that threat may not only be related directly to attitudes toward immigrants per se. Instead, it may also have an additional direct association with some of its specific measures. Such associations may prevent the attitudes scale to be invariant across countries. It must be noted, however, that we still could not explain all the between-group variance of the intercept of the item, thus suggesting that there still must be other non-identified contextual variables influencing the intercept's variation across countries.

In sum, the method demonstrates that even when measurement invariance is not supported by the data, it is still possible for researchers to try to understand why it is not given. Such an attempt may yield significant benefits in sociology: A full account of non-invariance by contextual variables may actually rectify non-invariance (Hox 2012, personal communication). The main challenge, however, for applying this type of approach is the selection of meaningful contextual predictors to explain non-invariance. In addition, a possible additional drawback of the approach is that it requires a sufficient amount of countries to identify the model and allow for a meaningful multilevel analysis (see, e.g., Meuleman and Billiet 2009). Nonetheless, incorporating MLSEM into future research will provide further opportunities to analyze large scale international surveys with a large number of countries and time points, and draw upon diverse sociological theories to identify sources of measurement non-invariance across various cultures or other groups of interest.

References

Ariely, Gal and Eldad Davidov. 2011. "Can We Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey." *Social Indicators Research* 104(2):271-86. doi: 10.1007/s11205-010-9693-5.

Billiet, Jaak. 2003. "Cross-Cultural Equivalence with Structural Equation Modeling." Pp. 247-64 in *Cross-Cultural Survey Methods*, edited by Janet A. Harkness, Fons J.R. van de Vijver, and Peter P. Mohler. Hoboken, NJ: Wiley.

Blalock, Hubert M. 1967. *Toward a Theory of Minority Group Relations.* New York: Wiley.

Blumer, Herbert. 1958. "Race Prejudice as a Sense of Group Position." *Pacific Sociological Review* 1(1):3-7.

Braun, Michael, Dorothée Behr, and Lars Kaczmirek. 2013. "Assessing Cross-National Equivalence of Measures of Xenophobia: Evidence from Probing in Web Surveys." *International Journal of Public Opinion Research* 25(3):383-95. doi: http://dx.doi.org/10.1093/ijpor/eds034.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables.* New York: Wiley.

Byrne, Barbara M. 2010. *Structural Equation Modeling With AMOS. Basic Concepts, Applications, and Programming*. 3rd ed. New York: Routledge.

Byrne, Barbara M., Richard J. Shavelson, and Bengt O. Muthén. 1989. "Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance." *Psychological Bulletin* 105(3):456-66.

Byrne, Barbara M. and Fons J.R. van de Vijver. 2010. "Testing for Measurement and Structural Equivalence in Large-Scale Cross-Cultural Studies: Addressing the Issue of Nonequivalence." *International Journal of Testing* 10:107-32. doi: 10.1080/15305051003637306.

Ceobanu, Alin M. and Xavier Escandell. 2011. “Paths to Citizenship? Public Views on the Extension of Rights to Legal and Second-Generation Immigrants in Europe.” *British Journal of Sociology* 62(2):221-40. doi: 10.1111/j.1468-4446.2011.01362.x.

Chen, Fang Fang. 2007. “Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance.” *Structural Equation Modeling* 14(3):464-504. doi: 10.1080/10705510701301834.

Chen, Gilad, Paul D. Bliese, and John E. Mathieu (2005): “Conceptual Framework and Statistical Procedures for Delineating and Testing Multilevel Theories of Homology.” *Organizational Research Methods* 8(4):375-409.

Cheung, Mike W.-L. and Kevin Au. 2005. “Applications of Multilevel Structural Equation Modeling to Cross-Cultural Research.” *Structural Equation Modeling* 12(4):598-619. doi: 10.1207/s15328007sem1204_5.

Cheung, Mike W.-L., Kwok Leung, and Kevin Au. 2006. “Evaluating Multilevel Models in Cross-Cultural Research: An Illustration with Social Axioms.” *Journal of Cross-Cultural Psychology* 37(5):522-41. doi: 10.1177/0022022106290476.

Cieciuch, Jan, Eldad Davidov, Peter Schmidt, René Algesheimer, and Shalom H. Schwartz. 2014. “Comparing Results of an Exact Versus an Approximate (Bayesian) Measurement Invariance Test: A Cross-Country Illustration with a Scale to Measure 19 Human Values.” *Frontiers in Psychology* 5:982. doi: 10.3389/fpsyg.2014.00982.

Davidov, Eldad. 2008. “A Cross-Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey.” *Survey Research Methods* 2:33-46.

Davidov, Eldad. 2009. “Measurement Equivalence of Nationalism and Constructive Patriotism in the ISSP: 34 Countries in a Comparative Perspective.” *Political Analysis* 17:64-82. doi: 10.1093/pan/mpn014.

Davidov, Eldad. 2011. “Nationalism and Constructive Patriotism: A Longitudinal Test of Comparability in 22 Countries with the ISSP.” *International Journal of Public Opinion Research* 23(1):88-103. doi: 10.1093/ijpor/edq031.

Davidov, Eldad, Hermann Dülmer, Elmar Schlüter, Peter Schmidt, and Bart Meuleman. (2012). “Using a Multilevel Structural Equation Modeling Approach to Explain Cross-Cultural Measurement Noninvariance.” *Journal of Cross-Cultural Psychology* 43: 558-75. doi: 10.1177/0022022112438397.

Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt, and Jaak Billiet. 2014. “Measurement Equivalence in Cross-National Research.” *Annual Review of Sociology* 40:55-75. doi: 10.1146/annurev-soc-071913-043137.

Davidov, Eldad, Peter Schmidt, and Jaak Billiet. eds. 2011. *Cross-Cultural Analysis: Methods and Applications.* New York: Routledge.

Davidov, Eldad, Peter Schmidt, and Shalom H. Schwartz. 2008. “Bringing Values Back In. The Adequacy of the European Social Survey to Measure Values in 20 Countries.” *Public Opinion Quarterly* 72:420-45. doi: 10.1093/poq/nfn035.

De Beuckelaer, Alain, Filip Lievens, and Gilbert Swinnen. 2007. “Measurement Equivalence in the Conduct of a Global Organizational Survey across Countries in Six Cultural Regions.” *Journal of Occupational and Organizational Psychology* 80:575-600. doi: 10.1348/096317907X173421.

De Beuckelaer, Alain and Gilbert Swinnen. 2011. “Biased Latent Variable Mean Comparisons due to Measurement Noninvariance: A Simulation Study.” Pp. 117-47 in *Cross-Cultural Analysis. Methods and Applications,* edited by Eldad Davidov, Peter Schmidt, and Jaak Billiet. New York: Routledge.

Fischer, Ronald, Taciano L. Milfont, and Valdiney V. Gouveia. 2011. “Does Social Context Affect Value Structures? Testing the Within-Country Stability of Value Structures

With a Functional Theory of Values." *Journal of Cross-Cultural Psychology* 42(2):253-70. doi: 10.1177/0022022110396888.

Fontaine, Johnny R.J. 2008. "Traditional and Multilevel Approaches in Cross-Cultural Research: An Integration of Methodological Frameworks." Pp. 65-92 in *Multilevel Analysis of Individuals and Cultures,* edited by Fons J.R. van de Vijver, Dianne A. van Hemert, and Ype H. Poortinga. New York: Lawrence Erlbaum Associates.

Fontaine, Johnny R.J., Ype H. Poortinga, Luc Delbeke, and Shalom H. Schwartz. 2008. "Structural Equivalence of the Values Domain Across Cultures: Distinguishing Sampling Fluctuations From Meaningful Variation." *Journal of Cross-Cultural Psychology* 39(4):345-65. doi: 10.1177/0022022108318112.

Gorodzeisky, Anastasia and Moshe Semyonov. 2009. "Terms of Exclusion: Public Views towards Admission and Allocation of Rights to Immigrants in European Countries." *Ethnic and Racial Studies* 32(3):401-23. doi: 10.1080/01419870802245851.

Hempel, Carl G. 1952. *Fundamentals of Concept Formation in Empirical Science (International Encyclopedia of Unified Science, vol. 2 no. 7).* Chicago: University of Chicago Press.

Hochman, Oshrat, Rebeca Raijman and Peter Schmidt. 2015. "National Identity and Exclusion of Non-Ethnic Migrants: Germany and Israel in Comparative Perspective. Pp. 64-82 in *Dynamics of National Identity,* edited by Jürgen Grimm, Leonie Huddy, Peter Schmidt and Josef Seethaler. London: Routledge.Horn, John L. and J. J. McArdle. 1992. "A Practical and Theoretical Guide to Measurement Invariance in Aging Research." *Experimental Aging Research* 18(3):117-44. doi: 10.1080/03610739208253916.

Hox, Joop. 2010. *Multilevel Analysis. Techniques and Applications.* Second Edition. New York, NJ: Routledge.

Hu, Li-tze, and Peter M. Bentler. 1999. “Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives.” *Structural Equation Modeling* 6(1):1-55. doi: 10.1080/10705519909540118.

ISSP Research Group. 2012. International Social Survey Programme: National Identity II - ISSP 2003. GESIS Data Archive, Cologne. ZA3910 Data file Version 2.1.0, doi:10.4232/1.11449.

Jak, Suzanne. 2014. “Testing Strong Factorial Invariance Using Three-Level Structural Equation Modeling.” *Frontiers in Psychology* 5:745. doi: 10.3389/fpsyg.2014.00745.

Jak, Suzanne, Frans J. Oort, and Conor V. Dolan. 2013. “A Test for Cluster Bias: Detecting Violations of Measurement Invariance across Clusters in Multilevel Data.” *Structural Equation Modeling* 20(2):265-82. doi: 10.1080/10705511.2013.769392.

Jak, Suzanne, Frans J. Oort, and Conor V. Dolan. 2014a. “Using Two-Level Factor Analysis to Test for Cluster Bias in Ordinal Data”. *Multivariate Behavioral Research* 49(6):544-553.

Jak, Suzanne, Frans J. Oort, and Conor V. Dolan. 2014b. “Measurement Bias in Multilevel Data.” *Structural Equation Modeling* 21(1):31-39. doi: 10.1080/10705511.2014.856694.

Jöreskog, Karl G. 1971. “Simultaneous Factor Analysis in Several Populations.” *Psychometrika* 36(4):409-26.

Kuha, Jouni, and Moustaki, Irini. 2015, July 27. “Nonequivalence of Measurement in Latent Variable Modeling of Multigroup Data: A Sensitivity Analysis”. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000031

Kim, Eun Sook, Myeongsun Yoon, Yao Wen, Wen Luo, and Oi-man Kwok (2015). “Within-Level Group Factorial Invariance with Multilevel Data: Multilevel Factor Mixture and Multilevel MIMIC Models”. *Structural Equation Modeling: A Multidisciplinary Journal,* 22(4):603-616.

Lee, Jaehoon, Todd D. Little, and Kristopher J. Preacher. 2011. “Methodological Issues in Using Structural Equation Models for Testing Differential Item Functioning.” Pp. 55-84 in *Cross-Cultural Analysis. Methods and Applications,* edited by Eldad Davidov, Peter Schmidt, and Jaak Billiet. New York: Routledge.

Levanon, Asaf, and Noah Lewin-Epstein. 2010. “Grounds for Citizenship: Public Attitudes in Comparative Perspective.” *Social Science Research* 39(3):419-31. doi: 10.1016/j.ssresearch.2009.12.001.

Little, Todd D. 2013. *Longitudinal Structural Equation Modeling.* New York: Guilford Press.

Little, Todd D., Ulman Lindenberger, and John R. Nesselroade. 1999. “On Selecting Indicators for Multivariate Measurement and Modeling With Latent Variables: When “Good” Indicators Are Bad and “Bad” Indicators Are Good.” *Psychological Methods* 4(2):192-211.

Marsh, Herbert W., Kit-Tai Hau, and Zhonglin Wen. 2004. „In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler’s (1999) Findings.” *Structural Equation Modeling* 11(3):320-41. doi: 10.1207/s15328007sem1103_2.

Meuleman, Bart and Jaak Billiet. 2009. “A Monte Carlo Sample Size Study: How Many Countries are Needed for Accurate Multilevel SEM?” *Survey Research Methods* 3(1):45-58.

Meuleman, Bart, Eldad Davidov, and Jaak Billiet. 2009. “Changing Attitudes toward Immigration in Europe, 2002-2007. A Dynamic Group Conflict Theory Approach.” *Social Science Research* 38:352-65. doi: 10.1016/j.ssresearch.2008.09.006.

Muthén, Bengt O. 1985. “A Method for Studying the Homogeneity of Test Items with Respect to Other Relevant Variables.” *Journal of Educational Statistics* 10(2):121-32. doi: 10.3102/10769986010002121.

Muthén, Bengt O. 1989. “Latent Variable Modeling in Heterogeneous Populations.” *Psychometrika* 54(4):557-85.

Muthén, Bengt O. 1991. “Multilevel Factor Analysis of Class and Student Achievement Components.” *Journal of Educational Measurement* 28(4):338-54.

Muthén, Bengt O. 1994. “Multilevel Covariance Structure Analysis.” *Sociological Methods & Research* 22(3):376-98. doi: 10.1177/0049124194022003006.

Muthén, Bengt O. 2011. *Mean and Covariance Structure Analysis of Hierarchical Data*. Department of Statistics, UCLA.

Muthén, Bengt O. and Tihomir Asparouhov. 2013. “BSEM Measurement Invariance Analysis.” *Mplus Web Note* No.17, Retrieved January 11, 2013. Available online at: www.statmodel.com.

Muthén, Bengt O., Chih-Fen Kao, and Leigh Burstein. 1991. “Instructionally Sensitive Psychometrics: Application of a New IRT-Based Detection Technique to Mathematics Achievement Test Items.” *Journal of Educational Measurement* 28(1):1-22.

Muthén, Linda K. and Bengt O. Muthén. 1998-2014. *Mplus User’s Guide.* Los Angeles: Muthén & Muthén.

Oberski, Daniel L. 2009. *Jrule for Mplus Version* 0.91 (beta) [Computer software], available at: https://github.com/daob/JruleMplus/wiki.

Quillian, Lincoln. 1995. “Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-Immigrant and Racial Prejudice in Europe.” *American Sociological Review* 60(4):586–611.

Rabe-Hesketh, Sophia, Skrondal, Anders, and Pickles, Andrew. 2004. “Generalized Multilevel Structural Equation Modeling.” *Psychometrika* 69(2):167-90.

Raijman, Rebeca, Eldad Davidov, Peter Schmidt, and Oshrat Hochman. 2008. “What Does a Nation Owe Non-Citizens? National Attachments, Perception of Threat and Attitudes towards Granting Citizenship Rights in a Comparative Perspective." *International Journal of Comparative Sociology* 49(2-3):195-220. doi: 10.1177/0020715208088912.

Rammstedt, Beatrice, Lewis R. Goldberg, and Ingwer Borg. 2010. "The Measurement Equivalence of Big Five Factor Markers for Persons with Different Levels of Education." *Journal for Research in Personality* 44(1):53-61. doi:10.1016/j.jrp.2009.10.005.

Reeskens, Tim and Marc Hooghe. 2010. "Beyond the Civic-Ethnic Dichotomy: Investigating the Structure of Citizenship Concepts across Thirty-Three Countries." *Nations and Nationalism* 16(4):579-97. doi: 10.1111/j.1469-8129.2010.00446.x.

Saris, Willem E., Albert Satorra, and William M. van der Veld. 2009. "Testing Structural Equation Models or Detection of Misspecifications?" *Structural Equation Modeling* 16(4): 561-82. doi: 10.1080/10705510903203433.

Scheepers, Peer, Mérove Gijsberts, and Marcel Coenders. 2002. "Ethnic Exclusionism in European Countries. Public Opposition to Civil Rights for Legal Migrants as a Response to Perceived Ethnic Threat." *European Sociological Review* 18(1):17-34.

Schlüter, Elmar and Bart Meuleman. 2009. *Measurement Inequivalence as a Source of Useful Information. Using Multi-level SEM to Explain Why Measurements are Inequivalent.* Paper presented at the 3rd ESRA conference, June 29-July 3, Warsaw, Poland.

Schlüter, Elmar and Peer Scheepers. 2010. "The Relationship between Outgroup Size and Anti-Outgroup Attitudes: A Theoretical Synthesis and Empirical Test of Group Threat- and Intergroup Contact Theory." *Social Science Research* 39(2):285-95. doi: 10.1016/j.ssresearch.2009.07.006.

Schneider, Silke L. 2008. "Anti-Immigrant Attitudes in Europe: Outgroup-Size and Perceived Ethnic Threat." *European Sociological Review* 24(1):53-67. doi: 10.1093/esr/jcm034.

Selig, James P., Noel A Card, and Todd D. Little. 2008. "Latent Variable Structural Equation Modeling in Cross-Cultural Research: Multigroup and Multilevel Approaches." Pp.

93-119 in *Multilevel Analysis of Individuals and Cultures*, edited by Fons J.R. van de Vijver, Dianne A. van Hemert, and Ype H. Poortinga. New York: Lawrence Erlbaum Associates.

Semyonov, Moshe, Rebeca Raijman, and Anastasia Gorodzeisky. 2006. "The Rise of Anti-Foreigner Sentiment in European Societies, 1988-2000." *American Sociological Review* 71(3):426-49.

Steenkamp, Jan-Benedict E.M. and Hans Baumgartner. 1998. "Assessing Measurement Invariance in Cross-National Consumer Research." *Journal of Consumer Research* 25(1):78-90. doi: 10.1086/209528.

Steinmetz, Holger. 2011. "Estimation and Comparison of Latent Means across Cultures. Pp. 85-116 in *Cross-Cultural Analysis. Methods and Applications,* edited by Eldad Davidov, Peter Schmidt, and Jaak Billiet. New York: Routledge.

Steinmetz, Holger. 2013. "Analyzing Observed Composite Differences across Groups: Is Partial Measurement Invariance Enough?" *Methodology* 9(1): 1-12. doi: 10.1027/1614-2241/a000049.

The World Bank Group. 2014. *World Development Indicators*. Retrieved December 3, 2014 (http://data.worldbank.org/indicator/SM.POP.TOTL.ZS/countries?page=3&display=defaut).

Thompson, Marilyn S. and Samuel B. Green. 2013. "Evaluating Between-Group Differences in Latent Variable Means." Pp. 163-218 in *Structural Equation Modeling: A Second Course,* 2nd ed., edited by Gregory R. Hancock and Ralph O. Mueller. Charlotte, NC: Information Age Publishing.

Vandenberg, Robert J. and Charles E. Lance. 2000. "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research." *Organizational Research Methods* 3(1):4-70. doi: 10.1177/109442810031002.

Van de Schoot, Rens, Anouck Kluytmans, Lars Tummers, Peter Lugtig, Joop Hox, and Bengt O. Muthén. 2013. "Facing Off with Scylla and Charybdis: A Comparison of Scalar, Partial, and the Novel Possibility of Approximate Measurement Invariance." *Frontiers in Psychology* 4:770. doi: 10.3389/fpsyg.2013.00770.

Van der Veld, William M. and Willem E. Saris. 2011. "Causes of Generalized Social Trust." Pp. 207-47 in *Cross-Cultural Analysis: Methods and Applications,* edited by Eldad Davidov, Peter Schmidt, and Jaak Billiet. New York: Routledge.

West, Stephen G., Aaron B. Taylor, and Wei Wu. 2012. "Model Fit and Model Selection in Structural Equation Modeling." Pp. 209-31 in *Handbook of Structural Equation Modeling,* edited by Rick H. Hoyle. New York: The Guilford Press.

Woods, Carol M. 2009. "Evaluation of MIMIC-Model Methods to DIF Testing with Comparison to Two-Group Analysis." *Multivariate Behavioral Research* 44(1):1-27. doi: 10.1080/00273170802620121.

Wright, Matthew. 2011. "Diversity and the Imagined Community: Immigrant Diversity and Conceptions of National Identity." *Political Psychology* 32(5):837-62. doi: 10.1111/j.1467-9221.2011.00843.x

Zercher, Florian, Peter Schmidt, Jan Cieciuch, and Eldad Davidov. 2015. "The Comparability of the Universalism Value over Time and Across Countries in the European Social Survey: Exact Versus Approximate Measurement Equivalence." *Frontiers in Psychology* 6:733. doi: 10.3389/fpsyg.2015.00733.

Table 1: Correlations, Variances, and Covariances for the Indicators Measuring Attitudes toward Granting Citizenship Rights to Immigrants

			Within and Between Countries - Correlations and Covariances		
			1	2	3
		within			
1	Jus solis (v59a)		1.120	*.378*	*.309*
2	Jus sanguinis (v60a)		.349	.763	*.265*
3	Jus domicile(v61a)		.386	.273	1.393
		between			
1	Jus solis (v59a)		.078	*.633*	*.344*
2	Jus sanguinis (v60a)		.038	.045	*.542*
3	Jus domicile(v61a)		.032	.039	.112

Source: ISSP data 2003

Note: Italic entries in the upper diagonal are the correlations, entries in the diagonal are variances, and entries in the lower diagonal are covariances; the total sample includes 38,830 respondents from 35 countries (two German samples: East and West, two samples from Israel: one Jewish and one Arab, Taiwan excluded)

Table 2: Multigroup CFA (MGCFA)

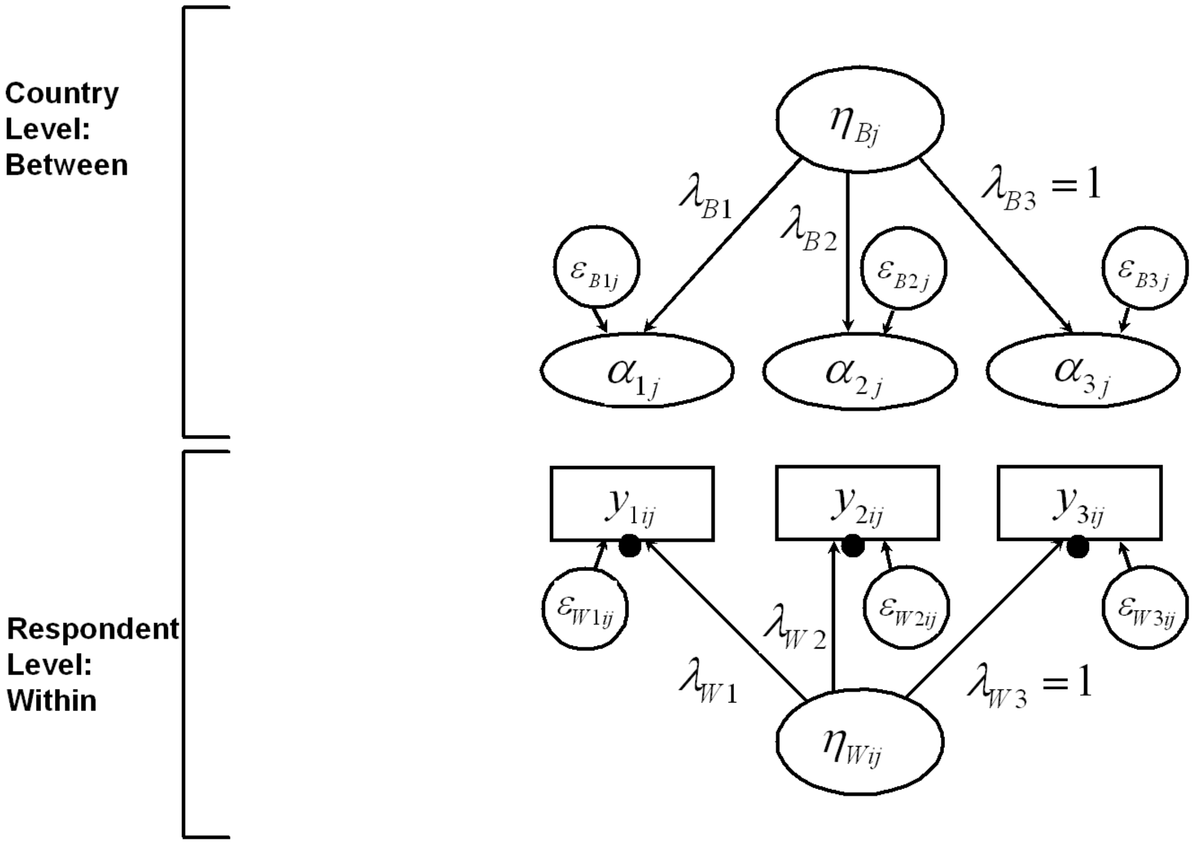
	χ^2	df	CFI	RMSEA	SRMR
Configural Invariance	0	0	1.00	.000 [.000 - .000]	.000
Metric Invariance	323.9	70	.979	.055 [.049 - .061]	.031
Scalar Invariance	4672.8	140	.627	.163 [.159 - .167]	.109

Table 3: Multilevel CFA and Multilevel SEM Models

	Model 1: 2 Level Model only with Level 1 Factor		Model 2: 2 Level CFA		Model 3: 2 Level SEM	
AIC	326859.352		326834.327		326824.747	
SRMR Within	.000		.000		.000	
SRMR Between	.368		.000		.021	
RMSEA	.015		.000		.000	
Respondent Level (Level 1) Country Level (Level 2)	38,830 Respondents 35 Countries		38,830 Respondents 35 Countries		38,830 Respondents 35 Countries	
Confirmatory Factor Analysis	b	z	b	z	b	z
Intercept Level 2						
Jus solis (v59a)	2.824	59.669**	2.824	59.286**	2.888	51.032**
Jus sanguinis (v60a)	2.975	82.599**	2.975	82.089**	3.053	61.291**
Jus domicile (v61a)	2.183	38.478**	2.183	38.291**	2.371	33.512**
Factor Loadings					b	z
Jus solis (v59a)			1.000	-.-	1.000	-.-
Jus sanguinis (v60a)			1.198	4.727**	1.201	5.554**
Jus domicile (v61a)			1.026	2.946**	.782	50.789*
Jus solis (v59a)	1.000	-.-	1.000	-.-	1.000	-.-
Jus sanguinis (v60a)	.709	50.040**	.709	50.038**	.709	50.058**
Jus domicile (v61a)	.782	50.728**	.782	50.726**	.782	50.789**
Variances/Residual Variances	Variance	z	Variance	z	Variance	z
Jus solis (v59a)	.077	4.123**	.047	4.085**	.047	4.085**
Jus sanguinis (v60a)	.045	4.094**	-.-	-.-	-.-	-.-
Jus domicile (v61a)	.111	4.130**	.079	4.107**	.064	4.081**
Jus solis (v59a)	.627	58.991**	.627	58.990**	.627	59.015**
Jus sanguinis (v60a)	.516	84.683**	.516	84.683**	.516	84.691**
Jus domicile (v61a)	1.091	111.356**	1.091	111.356**	1.091	111.380**
Regression	b	z	b	z	b	z
Predictor Jus domicile (v61a)						
Foreign Born (%)					-.015	-2.951**
Predictors Citizenship (betw.)						
Foreign Born (%)					-.007	-1.998*
Variance Comp./Residual Var. Level 2	Variance	z	Variance	z	Variance	z
Intercept Level 2: Citizenship (betw.)			.031	2.059*	.028	2.315*
Level 1: Citizenship (within)	.493	42.116**	.493	42.114**	.493	42.140**

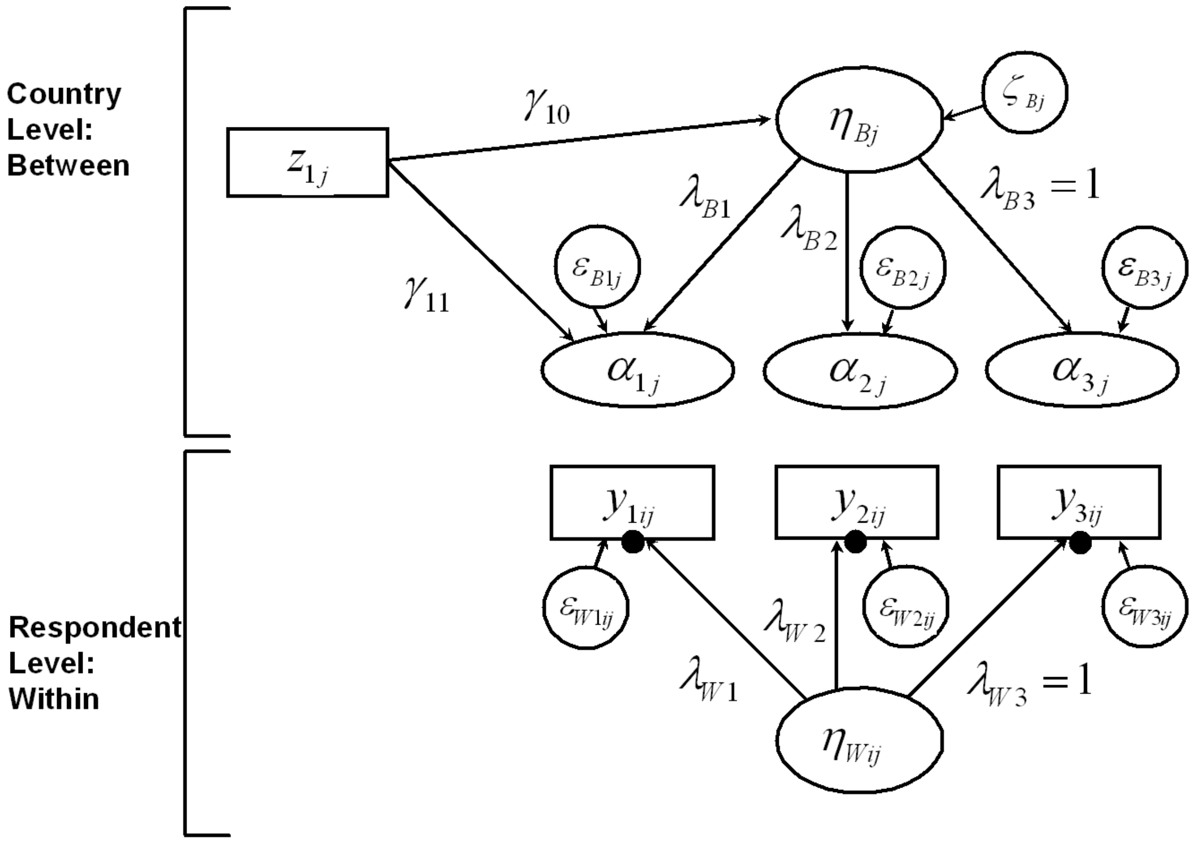
* $p \leq 0.05$; ** $p \leq 0.01$; Estimator: Full Maximum Likelihood (ML); for Model 2 the residual variance of jus sanguinis (v60a) at Level 2 was non-significant ($p = 0.996$) and has been fixed to zero for this reason. Since after controlling for the z-variable "Foreign Born" the factor loadings of jus domicile became nearly identical across both levels (0.782 for Level 1 and 0.776 for Level 2), an equality constraint was imposed on its factor loadings across levels (Model 3). The substantive implications do not change when the constraint is removed from the model. Level 2 predictors were tested one-tailed.

Figure 1a. A Two-Level CFA with Three Indicators



Note: Rectangles represent k=3 indicators on the within level; one-sided arrows represent causal effects; ellipses η_{Wij} and η_{Bj} represent the latent variable on the within and between levels, respectively; small circles next to the rectangles refer to the within-level error term ε_W for respondent i of country j on indicator variable k; black dots at the within-level refer to the country-specific intercepts of an indicator variable k which appear at the between level as large ellipses α_{kj} (often named the y-between); small circles next to the indicators on the between level refer to the between-level error term ε_B (random term). Since y_{kij} is predicted by η_{Wij}, the arrows from the latent within-level factor lead to the indicator variables and not to the black dots.

Figure 1b. A Two-Level SEM with Three Indicators and One Level 2 Predictor



Note: γ_{10} represents the regression coefficient of the latent variable η_{Bj} on the country-level predictor variable z_1, γ_{11} refers to the regression coefficient from the country-specific intercept α_l on the country-level predictor z_1 (this regression is intended to explain measurement invariance in the intercept of the indicator variable y_1), and ζ_{Bj} is the between-level residual error term of the latent variable η_{Bj}.

Figure 2a: A Two-Level Model with a One-Level CFA (Model 1)

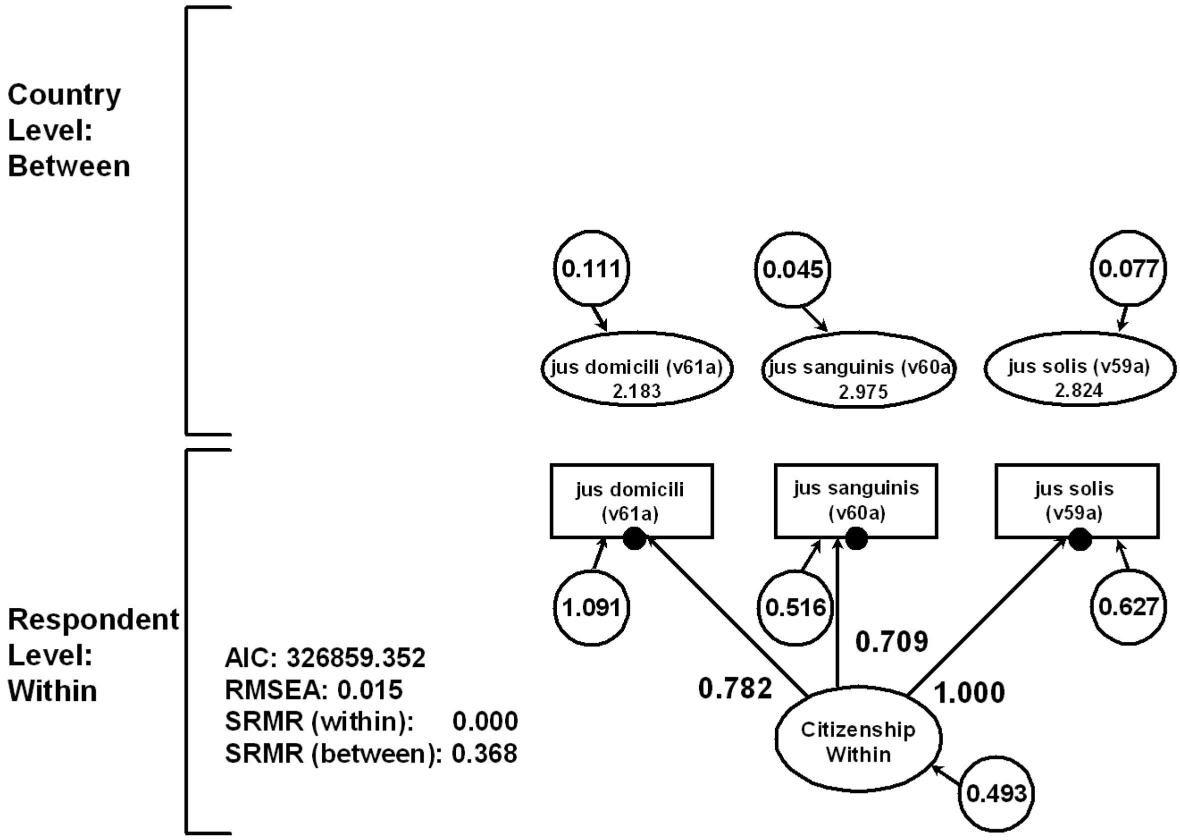


Figure 2b: A Two-Level CFA (Model 2)

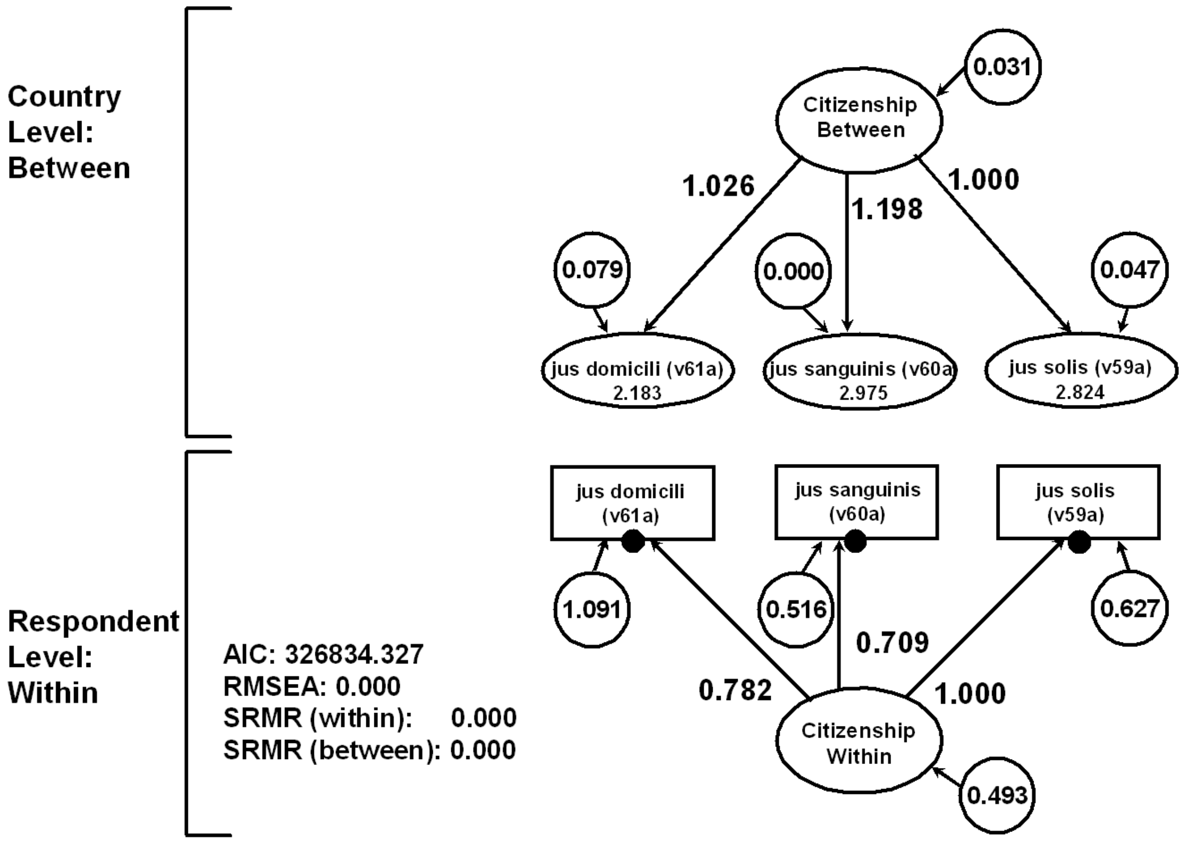
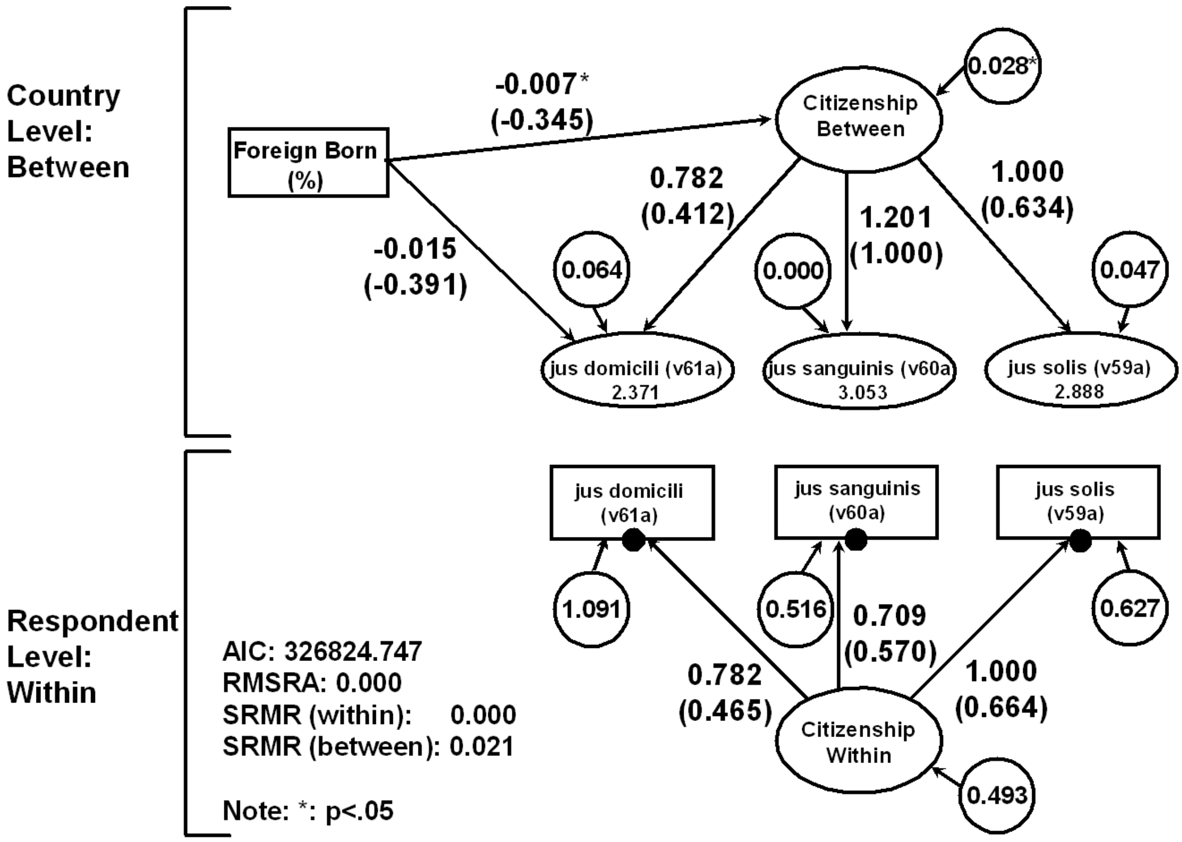


Figure 2c: A Two-Level SEM (Model 3)



Note: *p < 0.05, all other coefficients are significant at the 1% level (p < 0.01); standardized coefficients in parentheses. Factor loadings of "jus domicile" were constrained to be equal across both levels.

Appendix:

International Migrant Stock (% of Population): Percentage of Foreign-Born Immigrants in a Country Averaged Between 1995 and 1999 for the Countries (or Groups) in our Analysis

Australia (21.324), Austria (12.446), Bulgaria (0.554), Canada (17.194), Chile (0.941), Czech Republic (4.393), Denmark (5.683), Finland (2.023), France (10.220), Germany-East (11.009), Germany-West (11.009), Great Britain (7.223), Hungary (2.839), Ireland (7.307), Israel Arabs (34.613), Israel Jews (34.613), Japan (1.086), Korea (South, 1.296), Latvia (21.213), Netherlands (8.970), New Zealand (16.177), Norway (5.431), Philippines (0.302), Poland (2.497), Portugal (5.265), Russia (7.903), Slovakia (2.117), Slovenia (10.059), South Africa (2.806), Spain (2.643), Sweden (10.260), Switzerland (20.896), United States (10.711), Uruguay (2.891), Venezuela (4.615)

Source: The World Bank Group (2014)

Note: Taiwan excluded (no information about the international migrant stock). East and West Germany include the same number because the official statistics publishes the same number for both.